ON GENERALIZED CANONICAL CORRELATION ANALYSIS

Michel van de Velden
Econometric Institute, Erasmus University Rotterdam
P.O.Box 1738, 3000DR Rotterdam, The Netherlands
E-mail:vandevelden@ese.eur.nl

Introduction

In generalized canonical correlation analysis several sets of variables are analyzed simultaneously. This makes the method suited for the analysis of various types of data. For example, in marketing research, subjects may be asked to rate a set of objects on a set of attributes. For each individual, a data matrix can then be constructed where the objects are represented row-wise and the attributes column-wise. Then, using generalized canonical correlation analysis a graphical representation, sometimes referred to as a perceptual map, can be made on the basis of the individuals' observation matrices. Note that, the observation matrices do not necessarily contain the same attributes.

Several generalizations of canonical correlation analysis have been proposed. Some of these are discussed and compared in Kettenring (1971) and Gower (1989). In this paper, we shall concern ourselves with the generalization proposed by Carroll (1968). Carroll's approach has some attractive properties that makes the method well fit for the analysis of multiple-set data. First of all, computationally, the method is straightforward as its solution is based on an eigenequation. Secondly, the method is closely related to several well-known multivariate techniques. In particular, principal component analysis and multiple correspondence analysis. Thirdly, Carroll's generalization takes ordinary canonical correlation analysis as a special case. Although this last property is well known and already mentioned by Carroll (1968), a formal proof in the context of generalized canonical correlation analysis is not easy to find in the literature. Ten Berge (1979) does provide a proof in the context of factor rotation. In this paper, we will present a new proof of the equivalence. In addition, we propose a new generalized canonical correlation analysis approach that takes classical canonical correlationa analysis as a special case and always yields orthogonal canonical variates.

Canonical Correlation Analysis

In canonical correlation analysis (CCA), Hotelling (1936) the aim is to find linear combinations for two sets of variables in such a way that the correlation between the two linear combinations is maximal. Let: $\mathbf{X}_1 : n \times p_1$ and $\mathbf{X}_2 : n \times p_2$ denote centered and standardized data matrices. The idea is to construct $\mathbf{w} = \mathbf{X}_1\mathbf{a}_1$ and $\mathbf{z} = \mathbf{X}_2\mathbf{a}_2$ so that the correlation between $\mathbf{w}$ and $\mathbf{z}$ is maximal.

The vectors $\mathbf{w}$ and $\mathbf{z}$ are the canonical variates. These canonical variates are standardized: $\mathbf{w}'\mathbf{w} = \mathbf{z}'\mathbf{z} = 1$. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are often referred to as canonical weights. Instead of obtaining two canonical variates, additional variates may be constructed that are orthogonal with respect to the previous ones. Then: $\mathbf{W} = \mathbf{X}_1\mathbf{A}_1$ and $\mathbf{Z} = \mathbf{X}_2\mathbf{A}_2$ and $\mathbf{W}'\mathbf{W} = \mathbf{Z}'\mathbf{Z} = \mathbf{I}$. The canonical variates can then be obtained by using the following singular value decomposition:

$$\mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{R}_{12}\mathbf{R}_{22}^{-\frac{1}{2}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}',$$

where $\mathbf{R}_{11}$ denotes the correlation matrix for the first set of variables: $\mathbf{R}_{11} = \mathbf{X}_1'\mathbf{X}_1$, $\mathbf{R}_{22}$ is the correlation matrix for the second set: $\mathbf{R}_{22} = \mathbf{X}_2'\mathbf{X}_2$, and $\mathbf{R}_{12}$ gives the between sets correlation matrix: $\mathbf{R}_{12} = \mathbf{X}_1'\mathbf{X}_2$.

The canonical weights are calculated as

$$\mathbf{A}_1 = \mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U} \text{ and } \mathbf{A}_2 = \mathbf{R}_{22}^{-\frac{1}{2}}\mathbf{V}$$

so that

$$\mathbf{W}'\mathbf{W} = \mathbf{A}_1'\mathbf{R}_{11}\mathbf{A}_1 = \mathbf{U}'\mathbf{U} = \mathbf{I},$$

$$\mathbf{Z}'\mathbf{Z} = \mathbf{A}_2'\mathbf{R}_{22}\mathbf{A}_2 = \mathbf{V}'\mathbf{V} = \mathbf{I},$$

and

$$\mathbf{W}'\mathbf{Z} = \mathbf{Z}'\mathbf{W} = \mathbf{\Lambda}.$$

Generalized Canonical Correlation Analysis

Carroll (1968) introduced a generalization of canonical correlation analysis of two sets of variables to n sets of variables. From here on, we shall refer to this generalization as GCCA. Crucial in Carroll's method is the introduction of a so-called *group configuration* $\mathbf{Y}$. The linear combinations are chosen in such a way that the sum of squared correlations between them and the group configuration is at a maximum:

$$\begin{aligned} \max\ R^2 &= \sum_{i=1}^{n} r\left(\mathbf{y}, \mathbf{X}_i\mathbf{a}_i\right)^2, \\ s.t.\ \mathbf{y}'\mathbf{y} &= 1. \end{aligned}$$

An alternative and convenient way of expressing the GCCA objective is as follows:

$$\min_{\mathbf{A},\mathbf{Y}}\ \phi = \text{trace} \sum_{i=1}^{n} \left(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i\right)' \left(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i\right)$$

$$s.t.\ \mathbf{Y}'\mathbf{Y} = \mathbf{I}_k.$$

This formulation is sometimes refered to as the homogeneity formulation due to its close link with homogeneity analysis as described in Gifi (1990).

The group configuration matrix $\mathbf{Y}$ can be obtained using the eigenequation

$$\text{(1)} \qquad \left(\sum_{i=1}^{n} \mathbf{P}_i\right) \mathbf{Y} = \mathbf{Y}\mathbf{\Gamma}.$$

where

$$\mathbf{P}_i = \mathbf{X}_i \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'.$$

The matrices $\mathbf{A}_i$ can be calculated as

$$\mathbf{A}_i = \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'\mathbf{Y}.$$

Equivalence of GCCA and CCA when there are two sets of variables

Carroll's generalized canonical correlation analysis takes ordinary canonical correlation analysis as a special case. Although this property is well known and already mentioned by Carroll (1968), a formal proof in the context of GCCA is not easy to find in the literature. Here we will prove the property by showing that, when there are only two sets of variables, orthogonality of the group configuration implies orthogonality (with respect to the appropriate variance matrices) of the canonical variates.To show the relationship between CCA and GCCA for two sets of variables, we first show that the CCA solution is also a GCCA solution.

$CCA \Longrightarrow GCCA:$

Note that the group configuration used in GCCA, is not present in CCA. On the other hand, the canonical weights are restricted in CCA but not in GCCA. That is, in CCA the correlation is maximized under the restrictions: $\mathbf{A}_1'\mathbf{R}_{11}\mathbf{A}_1 = \mathbf{I}$ and $\mathbf{A}_2'\mathbf{R}_{22}\mathbf{A}_2 = \mathbf{I}$. In GCCA only the group configuration is constrained to be orthonormal.

Using the CCA solution $\mathbf{W}$ and $\mathbf{Z}$, and substituting their sum in eigenequation 1, we get

$$
\begin{aligned}
& (\mathbf{P}_1 + \mathbf{P}_2)(\mathbf{W} + \mathbf{Z}) \\
= & \left(\mathbf{X}_1\mathbf{R}_{11}^{-1}\mathbf{X}_1' + \mathbf{X}_2\mathbf{R}_{22}^{-1}\mathbf{X}_2'\right)(\mathbf{W} + \mathbf{Z}) \\
= & \mathbf{X}_1\mathbf{R}_{11}^{-1}\mathbf{X}_1'\mathbf{W} + \mathbf{X}_2\mathbf{R}_{22}^{-1}\mathbf{X}_2'\mathbf{W} \\
& + \mathbf{X}_1\mathbf{R}_{11}^{-1}\mathbf{X}_1'\mathbf{Z} + \mathbf{X}_2\mathbf{R}_{22}^{-1}\mathbf{X}_2'\mathbf{Z} \\
= & \mathbf{X}_1\mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U} + \mathbf{X}_2\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U} \\
& + \mathbf{X}_1\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-\frac{1}{2}}\mathbf{V} + \mathbf{X}_2\mathbf{R}_{22}^{-\frac{1}{2}}\mathbf{V} \\
= & \mathbf{W} + \mathbf{Z} + \mathbf{X}_2\mathbf{R}_{22}^{-\frac{1}{2}}\mathbf{V}\mathbf{\Lambda} + \mathbf{X}_1\mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Lambda} \\
= & (\mathbf{W} + \mathbf{Z})(\mathbf{I} + \mathbf{\Lambda}).
\end{aligned}
$$

Furthermore as $\mathbf{W}'\mathbf{W} = \mathbf{Z}'\mathbf{Z} = \mathbf{I}$ and $\mathbf{W}'\mathbf{Z} = \mathbf{Z}'\mathbf{W} = \mathbf{\Lambda}$ it follows that

$$(\mathbf{W} + \mathbf{Z})'(\mathbf{W} + \mathbf{Z}) = 2(\mathbf{I} + \mathbf{\Lambda}).$$

Hence,

$$\left(\sum_{i=1}^{2}\mathbf{P}_i\right)\mathbf{Y} = \mathbf{Y}\mathbf{\Gamma},$$

where $\mathbf{\Gamma} = \mathbf{I} + \mathbf{\Lambda}$, and $\mathbf{Y} = \frac{1}{\sqrt{2}}(\mathbf{W} + \mathbf{Z})(\mathbf{I} + \mathbf{\Lambda})^{-\frac{1}{2}}$ so that $\mathbf{Y}'\mathbf{Y} = \mathbf{I}$.

Now, $\mathbf{A}_{GCCA,i} = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{Y}$, so that

$$
\begin{aligned}
\sqrt{2}\mathbf{A}_{GCCA,1} &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{W} + \mathbf{Z})(\mathbf{I} + \mathbf{\Lambda})^{-\frac{1}{2}} \\
&= \left(\mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U} + \mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-\frac{1}{2}}\mathbf{V}\right)(\mathbf{I} + \mathbf{\Lambda})^{-\frac{1}{2}} \\
&= \left(\mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U} + \mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Lambda}\right)(\mathbf{I} + \mathbf{\Lambda})^{-\frac{1}{2}} \\
&= \mathbf{R}_{11}^{-\frac{1}{2}}\mathbf{U}(\mathbf{I} + \mathbf{\Lambda})^{\frac{1}{2}} \\
&= \mathbf{A}_1(\mathbf{I} + \mathbf{\Lambda})^{\frac{1}{2}}.
\end{aligned}
$$

Similarly, $\sqrt{2}\mathbf{A}_{GCCA,2} = \mathbf{A}_2(\mathbf{I} + \mathbf{\Lambda})^{\frac{1}{2}}$.

Hence, the only difference between the methods concerns the scaling. In GCCA

$$\mathbf{A}_1'\mathbf{R}_{11}\mathbf{A}_1 = \frac{1}{2}(\mathbf{I} + \mathbf{\Lambda}) = \frac{1}{2}\mathbf{\Gamma}.$$

So, starting with a CCA solution, we have shown that the GCCA solution is equivalent. The other way around is less straightforward. It implies that the orthogonality constraint for the group configuration $\mathbf{Y}$ leads to orthogonality of the canonical variates.

$GCCA \Longrightarrow CCA:$

We have:

$$(2) \quad (\mathbf{P}_1 + \mathbf{P}_2)\mathbf{Y} = \mathbf{Y}\mathbf{\Gamma}.$$

Premultiplication by $\mathbf{P}_1$ and $\mathbf{P}_2$ yields

$$\mathbf{P}_1\mathbf{P}_2\mathbf{Y} = \mathbf{P}_1\mathbf{Y}\left(\mathbf{\Gamma} - \mathbf{I}\right) \tag{3}$$

and

$$\mathbf{P}_2\mathbf{P}_1\mathbf{Y} = \mathbf{P}_2\mathbf{Y}\left(\mathbf{\Gamma} - \mathbf{I}\right). \tag{4}$$

Hence

$$\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{Y} = \mathbf{P}_1\mathbf{Y}\mathbf{D}, \tag{5}$$

and

$$\mathbf{P}_2\mathbf{P}_1\mathbf{P}_2\mathbf{Y} = \mathbf{P}_2\mathbf{Y}\mathbf{D}, \tag{6}$$

where

$$\mathbf{D} = \left(\mathbf{\Gamma} - \mathbf{I}\right)^2.$$

It follows that

$$\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{Y}\mathbf{D}^{-1} = \mathbf{P}_1\mathbf{Y}.$$

Taking the transpose we see that

$$\mathbf{Y}'\mathbf{P}_1 = \mathbf{D}^{-1}\mathbf{Y}\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1$$

hence

$$\mathbf{Y}'\mathbf{P}_1\mathbf{Y} = \mathbf{D}^{-1}\mathbf{Y}'\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{Y}. \tag{7}$$

Furthermore, pre-multiplying (5) by $\mathbf{Y}'$ yields

$$\mathbf{Y}'\mathbf{P}_1\mathbf{Y} = \mathbf{Y}'\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{Y}\mathbf{D}^{-1}. \tag{8}$$

From (7) and (8) it follows that

$$\mathbf{D}\mathbf{Y}'\mathbf{P}_1\mathbf{Y} = \mathbf{Y}'\mathbf{P}_1\mathbf{Y}\mathbf{D},$$

hence, assuming $d_i \neq d_j$ (for $i \neq j$), it immediately follows from the diagonality of $\mathbf{D}$, that

$$\mathbf{Y}'\mathbf{P}_1\mathbf{Y} = \mathbf{D}_1,$$

where $\mathbf{D}_1$ is a diagonal matrix. Hence,

$$\mathbf{V}_1 = \mathbf{Y}\mathbf{P}_1\mathbf{D}_1^{-\frac{1}{2}}$$

is an orthonormal matrix.

In a similar fashion we obtain

$$\mathbf{Y}'\mathbf{P}_2\mathbf{Y} = \mathbf{D}_2,$$

where $\mathbf{D}_2$ is a diagonal matrix so that

$$\mathbf{V}_2 = \mathbf{Y}\mathbf{P}_2\mathbf{D}_2^{-\frac{1}{2}}$$

is orthonormal.

For the sake of completeness, let us derive $\mathbf{D}_1$ and $\mathbf{D}_2$. Pre-multiplication of (5) and (6) by $\mathbf{Y}'$ gives

$$\mathbf{Y}'\mathbf{P}_1\mathbf{P}_2\mathbf{P}_1\mathbf{Y} = \mathbf{Y}'\mathbf{P}_1\mathbf{Y}\mathbf{D} = \mathbf{D}_1\mathbf{D},$$

and

$$\mathbf{Y}'\mathbf{P}_2\mathbf{P}_1\mathbf{P}_2\mathbf{Y} = \mathbf{Y}'\mathbf{P}_2\mathbf{Y}\mathbf{D} = \mathbf{D}_2\mathbf{D}.$$

Using (3) and (4) these become

$$\mathbf{Y}'\mathbf{P}_1\mathbf{P}_2\mathbf{Y}\mathbf{D}^{\frac{1}{2}} = \mathbf{D}_1\mathbf{D},$$

and

$$\mathbf{Y}'\mathbf{P}_2\mathbf{P}_1\mathbf{Y}\mathbf{D}^{\frac{1}{2}} = \mathbf{D}_2\mathbf{D}.$$

Hence,

$$\mathbf{Y}'\mathbf{P}_1\mathbf{P}_2\mathbf{Y} = \mathbf{D}_1\mathbf{D}^{\frac{1}{2}}$$

and

$$\mathbf{Y}'\mathbf{P}_2\mathbf{P}_1\mathbf{Y} = \mathbf{D}_2\mathbf{D}^{\frac{1}{2}}.$$

Clearly, the diagonal matrices $\mathbf{D}_1\mathbf{D}^{\frac{1}{2}}$ and $\mathbf{D}_2\mathbf{D}^{\frac{1}{2}}$ are symmetric, so that

$$\mathbf{Y}'\mathbf{P}_1\mathbf{P}_2\mathbf{Y} = \mathbf{Y}'\mathbf{P}_2\mathbf{P}_1\mathbf{Y}$$

and

$$\mathbf{D}_1 = \mathbf{D}_2.$$

From (2) and the orthonormality of $\mathbf{Y}$ it follows that

$$\mathbf{Y}'\left(\mathbf{P}_1 + \mathbf{P}_2\right)\mathbf{Y} = \mathbf{Y}'\mathbf{P}_1\mathbf{Y} + \mathbf{Y}'\mathbf{P}_2\mathbf{Y} = \boldsymbol{\Gamma}.$$

Hence,

$$\boldsymbol{\Gamma} = \mathbf{D}_1 + \mathbf{D}_2 = 2\mathbf{D}_1 \Longrightarrow \mathbf{D}_1 = \mathbf{D}_2 = \frac{1}{2}\boldsymbol{\Gamma}.$$

Orthogonal Generalized Canonical Correlation Analysis

Recall the GCCA objective:

$$\text{(9)} \quad \min_{\mathbf{Y},\mathbf{A}} \operatorname{trace} \sum_{i=1}^{n} (\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)'(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)$$

$$s.t. \mathbf{Y}'\mathbf{Y} = \mathbf{I}.$$

In general, the canonical variates $\mathbf{X}_i\mathbf{A}_i$ will not be orthogonal. However, for the two variable case we saw that the orthogonality does hold. It also holds for the special case where the data matrices are so-called indicator matrices (i.e. matrices with in each row exactly one element equal to one and all other elements zero. This special case is equivalent to multiple correspondence analysis (MCA). Carroll (1968) notes that his method does not yield orthogonal canonical variates and suggest that this can be resolved by applying some orthogonalization step to the solution. Alternatively, we can formulate the problem and explicitly impose the restrictions. Hence,

$$\min_{\mathbf{Y},\mathbf{A}} \operatorname{trace} \sum_i (\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)'(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)$$

$$\begin{aligned} s.t.\ \mathbf{A}_i'\mathbf{A}_i &= \mathbf{I}.\ \text{(for all } i\text{) and} \\ \mathbf{Y}'\mathbf{Y} &= \mathbf{I}. \end{aligned}$$

To solve this problem we first consider:

$$\min_{\mathbf{B}} \operatorname{trace} \sum_i (\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)'(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)$$

$$s.t. \mathbf{A}_i'\mathbf{A}_i = \mathbf{I}.\ \text{(voor alle } i\text{)}.$$

That is, we assume the group configuration to be known.

The objective can be rewritten as

$$\begin{aligned} \min_{\mathbf{A}_i} \quad \phi &= n \operatorname{trace} \mathbf{Y}'\mathbf{Y} - 2\sum_i \operatorname{trace} \mathbf{Y}'\mathbf{X}_i\mathbf{A}_i \\ &\quad + \sum_i \operatorname{trace} \mathbf{A}_i'\mathbf{X}_i'\mathbf{X}_i\mathbf{A}_i. \end{aligned}$$

Setting up the Lagrangean we get

$$\begin{aligned} & n \operatorname{trace} \mathbf{Y}'\mathbf{Y} - 2\sum_i \operatorname{trace} \mathbf{Y}'\mathbf{X}_i\mathbf{A}_i + \\ & \sum_i \operatorname{trace} \mathbf{A}_i'\mathbf{X}_i'\mathbf{X}_i\mathbf{A}_i - \sum_i \operatorname{trace} \mathbf{L}_i(\mathbf{A}_i'\mathbf{A}_i - \mathbf{I}). \end{aligned}$$

Taking derivatives with respect to $\mathbf{A}_i$ yields

$$\begin{aligned} & -2 \operatorname{trace} \mathbf{Y}'\mathbf{X}_i d\mathbf{A}_i + 2 \operatorname{trace} \mathbf{A}_i'\mathbf{X}_i'\mathbf{X}_i d\mathbf{A}_i \\ & -2 \operatorname{trace} \mathbf{L}_i\mathbf{A}_i' d\mathbf{A}_i. \end{aligned}$$

Hence, the first order conditions for $\mathbf{A}_i$ are

$$\mathbf{A}_i'\mathbf{X}_i'\mathbf{X}_i - \mathbf{Y}'\mathbf{X}_i = \mathbf{L}_i\mathbf{A}_i'.$$

Postmultiplying by $\mathbf{A}_i$ and using the constraint gives

$$\mathbf{A}_i'\mathbf{X}_i'\mathbf{X}_i\mathbf{A}_i - \mathbf{Y}'\mathbf{X}_i\mathbf{A}_i = \mathbf{L}_i.$$

As the restrictions are symmetric, $\mathbf{L}_i$ is symmetric. Hence,

$$\mathbf{Y}'\mathbf{X}_i\mathbf{A}_i = \mathbf{A}_i'\mathbf{X}_i'\mathbf{Y}.$$

Then

$$\begin{aligned} (10)\quad \left(\mathbf{Y}'\mathbf{X}_i\mathbf{A}_i\right)^2 &= \mathbf{Y}'\mathbf{X}_i\mathbf{A}_i\mathbf{A}_i'\mathbf{X}_i'\mathbf{Y} \Rightarrow \\ (11)\quad \mathbf{Y}'\mathbf{X}_i\mathbf{A}_i &= \left(\mathbf{Y}'\mathbf{X}_i\mathbf{A}_i\mathbf{A}_i'\mathbf{X}_i'\mathbf{Y}\right)^{\frac{1}{2}}. \end{aligned}$$

Given $\mathbf{Y}$ and $\mathbf{X}_i$ a solution for $\mathbf{A}_i$ can be obtained using the singular value decomposition of $\mathbf{Y}'\mathbf{X}_i$. Let

$$\mathbf{Y}'\mathbf{X}_i = \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{V}_i',$$

where $\mathbf{U}_i$ and $\mathbf{V}_i$ are semi-orthogonal and $\mathbf{\Lambda}_i$ is a diagonal matrix with as diagonal elements the singular values of $\mathbf{Y}'\mathbf{X}_i$. It is then easily verified that for $\mathbf{A}_i = \mathbf{V}_i\mathbf{U}_i'$, (10) is satisfied.

Now, instead of assuming $\mathbf{Y}$ to be fixed, we can solve the objective for $\mathbf{Y}$ assuming the $\mathbf{A}_i's$ known:

$$\min_{\mathbf{Y}} \text{trace} \sum_i (\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)'(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i)$$

$$s.t. \mathbf{Y}'\mathbf{Y} = \mathbf{I}. \text{ (voor alle } i).$$

The objective can be rewritten as

$$\begin{aligned}\min_{\mathbf{Y}} \quad \phi \quad &= \quad n\,\text{trace}\,\mathbf{Y}'\mathbf{Y} - 2\sum_i \text{trace}\,\mathbf{Y}'\mathbf{X}_i\mathbf{A}_i \\ &\quad + \sum_i \text{trace}\,\mathbf{A}_i'\mathbf{X}_i'\mathbf{X}_i\mathbf{A}_i.\end{aligned}$$

or, as the terms involving only $\mathbf{X}_i$ and $\mathbf{A}_i$ play no role and the restriction implies that $n\,\text{trace}\,\mathbf{Y}'\mathbf{Y}$ is constant,:

$$\max_{\mathbf{Y}} 2\sum_i \text{trace}\,\mathbf{Y}'\mathbf{X}_i\mathbf{A}_i$$

$$s.t. \mathbf{Y}'\mathbf{Y} = \mathbf{I}.$$

Setting up the Lagrangean we get

$$2\sum_i \text{trace}\,\mathbf{Y}'\mathbf{X}_i\mathbf{A}_i - \text{trace}\,\mathbf{L}(\mathbf{Y}'\mathbf{Y} - \mathbf{I}).$$

Taking derivatives with respect $\mathbf{Y}$ yields and

$$2\sum_i \text{trace}\,\mathbf{A}_i{}'\mathbf{X}_i'd\mathbf{Y} - 2\,\text{trace}\,\mathbf{L}\mathbf{Y}'d\mathbf{Y} = 0.$$

Hence, the first order condition for $\mathbf{Y}$ is

$$\sum_i \mathbf{A}_i'\mathbf{X}_i' = \mathbf{L}\mathbf{Y}'.$$

Postmultiplication with $\mathbf{Y}$ yields

$$\sum_i \mathbf{A}_i'\mathbf{X}_i'\mathbf{Y} = \mathbf{L}$$

and as the restrictions are symmetric, $\mathbf{L}$ is symmetric so that

$$\left(\sum_i \mathbf{A}_i'\mathbf{X}_i'\right)\mathbf{Y} = \mathbf{Y}'\left(\sum_i \mathbf{X}_i\mathbf{A}_i\right).$$

Let

$$\sum_i \mathbf{X}_i\mathbf{A}_i = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'.$$

Then it is easily verified that a solution for $\mathbf{Y}$ can be obtained by considering

$$\mathbf{Y} = \mathbf{U}\mathbf{V}'.$$

Where $\mathbf{U}$ and $\mathbf{V}$ are the singular vectors corresponding to the largest singular values.

Iterating between these two problems will lead the objective to monotonically decrease. However, it is not sure that the thus obtained minimum is a global minimum. Therefore, it is recommended to repeat the procedure using different initial settings.

Note that, for convenience, the canonical weight matrices were constrained to be orthonormal here. For the canonical variates to be orthonormal, a rescaling must be carried out. That is,

$$\mathbf{A}_{rescaled,i} = \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-\frac{1}{2}} \mathbf{A}_i$$

so that

$$\mathbf{A}'_{rescaled,i}\mathbf{X}_i'\mathbf{X}_i\mathbf{A}_{rescaled,i} = \mathbf{I}.$$

Obviously, this method takes the original CCA approach as a special case.

Conclusion

In this paper, we presented a new prove for the equivalence between classical canonical correlation analysis and Carroll's (1968) generalized canonical correlation method (when only two sets of variables are considered). In addition, a new method was proposed that yields an orthogonal group configuration as well as orthogonal canonical variates. The new method is easy to calculate and takes classical canonical correlation analysis as a special case.

REFERENCES

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. Proceedings of the American Psychological Association, pages 227-228.

Gifi, A. (1990). Nonlinear Multivariate Analysis. Wiley, Chichester.

Gower, J.C. (1989). Generalized canonical analysis. In: R. Coppi and S. Bolasco, Editors, Multiway Data Analysis, North Holland, Amsterdam (1989), pp. 221–232.

Hotelling, H. (1936). Relations between two sets of variates. Biometrika, 28:321–377.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. Biometrika, 58:433-451.

Ten Berge, J.M.F. (1979). On the equivalence of two congruence rotation methods and orthogonal approximations. Psychometrika , 44, 359-363. (1973).